On this page ⌄

Glossary

Community Cloud

GPU instances connect individual compute providers to consumers through a vetted, secure peer-to-peer system.

Datacenter

A data center is a secure location where RunPod's cloud computing services, such as Secure Cloud and GPU Instances, are hosted. These data centers are equipped with redundancy and data backups to ensure the safety and reliability of your data.

Endpoint

An Endpoint refers to a specific URL where your serverless applications or services can be accessed. These endpoints provide standard functionality for submitting jobs and retrieving the output from job requests.

GPU Instance

GPU Instance is a container-based GPU instance that you can deploy. These instances spin up in seconds using both public and private repositories. They are available in two different types: Secure Cloud and Community Cloud.

Handler

A Handler is a function that is responsible for processing submitted inputs and generating the resulting output.

Ask AI

RunPod

RunPod is a cloud computing platform primarily designed for AI and machine learning applications.

Secure Cloud

GPU instances that run in T3/T4 data centers, providing high reliability and security.

Serverless GPU

Serverless GPU is a pay-per-second serverless GPU computing solution. It is designed to bring autoscaling to your production environment, meaning it can dynamically adjust computational resources based on your application's needs.

Template

A RunPod template is a Docker container image paired with a configuration.

✏️ Edit this page

Docs

Overview

Tutorials

AI APIs

Community

Discord ⬈

Ask AI

Contact us 🡕

More

Blog 🡕

GitHub 🡕

Ask AI